



The glycosyltransferase repertoire of the spikemoss *Selaginella moellendorffii* and a comparative study of its cell wall

Harholt, Jesper; Sørensen, Iben; Fangel, Jonatan Ulrik; Roberts, Alison; Willats, William George Tycho; Scheller, Henrik Vibe; Petersen, Bent L; Banks, Jo Ann; Ulvskov, Peter

Published in:
P L o S One

DOI:
[10.1371/journal.pone.0035846](https://doi.org/10.1371/journal.pone.0035846)

Publication date:
2012

Document version
Også kaldet Forlagets PDF

Citation for published version (APA):
Harholt, J., Sørensen, I., Fangel, J. U., Roberts, A., Willats, W. G. T., Scheller, H. V., Petersen, B. L., Banks, J. A., & Ulvskov, P. (2012). The glycosyltransferase repertoire of the spikemoss *Selaginella moellendorffii* and a comparative study of its cell wall. *P L o S One*, 7(5). <https://doi.org/10.1371/journal.pone.0035846>

The Glycosyltransferase Repertoire of the Spikemoss *Selaginella moellendorffii* and a Comparative Study of Its Cell Wall

Jesper Harholt¹, Iben Sørensen^{1,2}, Jonatan Fangel¹, Alison Roberts³, William G. T. Willats¹, Henrik Vibe Scheller^{4,5,6}, Bent Larsen Petersen¹, Jo Ann Banks⁷, Peter Ulvskov^{1*}

1 Department of Plant Biology and Biotechnology, VKR Research Centre “Pro-Active Plants,” University of Copenhagen, Frederiksberg, Denmark, **2** Department of Plant Biology, Cornell University, Ithaca, New York, United States of America, **3** Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island, United States of America, **4** Feedstocks Division, Joint Bioenergy Institute, Emeryville, California, United States of America, **5** Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America, **6** Department of Plant and Microbial Biology, University of California, Berkeley, California, United States of America, **7** Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana, United States of America

Abstract

Spike mosses are among the most basal vascular plants, and one species, *Selaginella moellendorffii*, was recently selected for full genome sequencing by the Joint Genome Institute (JGI). Glycosyltransferases (GTs) are involved in many aspects of a plant life, including cell wall biosynthesis, protein glycosylation, primary and secondary metabolism. Here, we present a comparative study of the *S. moellendorffii* genome across 92 GT families and an additional family (DUF266) likely to include GTs. The study encompasses the moss *Physcomitrella patens*, a non-vascular land plant, while rice and Arabidopsis represent commelinid and non-commelinid seed plants. Analysis of the subset of GT-families particularly relevant to cell wall polysaccharide biosynthesis was complemented by a detailed analysis of *S. moellendorffii* cell walls. The *S. moellendorffii* cell wall contains many of the same components as seed plant cell walls, but appears to differ somewhat in its detailed architecture. The *S. moellendorffii* genome encodes fewer GTs (287 GTs including DUF266s) than the reference genomes. In a few families, notably GT51 and GT78, *S. moellendorffii* GTs have no higher plant orthologs, but in most families *S. moellendorffii* GTs have clear orthologies with Arabidopsis and rice. A gene naming convention of GTs is proposed which takes orthologies and GT-family membership into account. The evolutionary significance of apparently modern and ancient traits in *S. moellendorffii* is discussed, as is its use as a reference organism for functional annotation of GTs.

Citation: Harholt J, Sørensen I, Fangel J, Roberts A, Willats WGT, et al. (2012) The Glycosyltransferase Repertoire of the Spikemoss *Selaginella moellendorffii* and a Comparative Study of Its Cell Wall. PLoS ONE 7(5): e35846. doi:10.1371/journal.pone.0035846

Editor: Edward Newbigin, University of Melbourne, Australia

Received: June 28, 2011; **Accepted:** March 26, 2012; **Published:** May 2, 2012

Copyright: © 2012 Harholt et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Villum-Kann Rasmussen grant to the Pro-Active Plant Centre (www.proactiveplants.life.ku.dk), by the Danish Research Council (FTP-09-066624 to PU), by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DEAC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy and by Agriculture and Food Research Initiative Competitive grant no. 2007-35318-18389 from the USDA National Institute of Food and Agriculture. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ulvskov@life.ku.dk

Introduction

Selaginella moellendorffii is an extant member of the lycopsid lineage, which diverged from the Euphyllophyta (ferns and seed plants) ca. 400 Myr ago [1]. *S. moellendorffii* was selected for sequencing by the Joint Genome Institute (JGI) because of its important evolutionary position among land plants and its small genome size [2,3]. Unlike bryophytes, which have a dominant gametophyte phase, the sporophyte is dominant phase of the lycopsid life cycle. Lycopside are further set apart from the mosses by the presence of a true vascular system for water and assimilate transport.

The moss *Physcomitrella patens* is the only non-vascular terrestrial plant and the only other non-angiosperm plant that has had its genome sequenced, and it is thus particularly important to the comparative genomic study of *S. moellendorffii* [4]. Bryophytes represent the oldest extant terrestrial plants and have diverged markedly from their Charophycean algal ancestors. Biochemical

studies of basal plants and Charophycean algae indicate that the widespread utilization of some cell wall features such as rhamnogalacturonan II (RG-II) and lignin, or at least uncon- densed monolignols, occurred concurrently with the colonization of land [5]. Xylan has been found throughout the green plant lineage while the common ancestor of all plants with xyloglucan (XyG) probably was a Charophycean alga [6]. XyG fine structure displays variations within vascular plants that appear to reflect taxonomy [7].

Presumably, specific elaborations in cell wall architecture are correlated with the particular lifestyles of individual plant species and in some cases mark major evolutionary transitions. One of the most important innovations following colonization of dry land was the tracheary element from which the term tracheophyte (vascular plant) is derived. Tracheary elements are hollow, rigid cells, which have undergone cell death and are specialized for water transport. They first appear in the fossil record during the late Silurian

period [8] and were important for the radiation of vascular plants during the Devonian period [9].

The complex polysaccharides of cell walls are synthesized by the coordinated activity of a large set of carbohydrate active enzymes, including glycosyltransferases (GTs) that catalyze the formation of glycosidic bonds between donor sugars and acceptor molecules and that are the primary focus of this study. The Carbohydrate Active enZyme, or CAZy, database [10] is of increasing importance to cell wall biology and we have used a CAZy-driven approach for analyzing the *S. moellendorffii* genome.

Some clades of putative cell wall biosynthetic enzymes are shared by *P. patens*, *S. moellendorffii* and seed plants whereas others diversified more recently and are restricted to one of the three lineages. These observations lead us to propose that although lycophytes share many cell wall features with the other major clades of land plants, *S. moellendorffii* (as well as *P. patens*) has undergone independent diversification of some gene families and retained ancestral genes that have been lost in seed plants. The evolutionary divergence of GTs and cell wall structure is presumably a reflection of the different adaptive strategies that predominate within the major lineages.

Comparative studies of cell wall biosynthesis proteomes should ideally be paralleled by comparative studies of the corresponding cell wall composition and architecture, in this case the analysis of the cell walls of *S. moellendorffii* compared to those of *P. patens* and of the fully sequenced angiosperms *Arabidopsis* and rice. In this regard, we present here biochemical and immunolocalisation analyses of *S. moellendorffii* cell walls as a foundation for discussing the cell wall biosynthetic proteome.

Results and Discussion

S. moellendorffii cell wall composition and architecture

Comprehensive Microarray Polymer Profiling. Cell wall polysaccharides were analysed using the microarray/antibody based Comprehensive Microarray Polymer Profiling or CoMPP technique [11] to obtain semi-quantitative information about the relative abundance of cell wall polysaccharides. CoMPP data for *P. patens* has previously been reported [11]. CoMPP analysis of *S. moellendorffii* detected many of the polysaccharides that are typically present in higher plants, including *Arabidopsis* (Figure 1). We found evidence of glycan epitopes that are typically associated with cellulose, pectins, xyloglucans, xylans, mannans, arabinogalactan-proteins (AGPs) and extensins. The XyG epitopes recognized by LM15 and CCRC-M1 (antibodies that recognize non-fucosylated and fucosylated XyG, respectively) were both present in *S. moellendorffii*. In contrast, fucosylated XyG epitope recognized by CCRC-M1 was essentially absent from *P. patens* [11,12]. The presence of mannan epitopes or polymers in the cell walls of *P. patens* and the Charophycean alga *Chara corallina* [11,13,14] in addition to *S. moellendorffii* indicates that mannose-containing cell wall polymers most likely evolved prior to the colonization of land. As with higher plants, the hemicellulosic and crystalline cellulose epitopes were detected almost exclusively in NaOH extracts. Extensin and AGP epitopes showed similar extraction patterns as observed for *Arabidopsis*.

Two surprising observations were made in the CoMPP data. Certain pectin epitopes, including homogalacturonan (HG, recognized by mAb JIM5), β -(1,4)-galactan (recognized by mAb LM5) and α -(1,5)-arabinan (recognized by mAb LM6) were extracted effectively only by NaOH and not by CDTA. This is in contrast to most other land plants, including *P. patens*, where these pectic polymers are more readily extracted with CDTA. The reduced CDTA solubility of *S. moellendorffii* HG may indicate a

different association between the pectic matrix and other cell wall components. This trend was confirmed in CoMPP analysis of another *Selaginella* species, *S. kraussiana* (data not shown). Moreover, a related pattern was observed in another basal land plant, the horsetail *Equisetum arvense*, where the JIM5 epitope in certain tissues was not completely extractable with CDTA but in the subsequent NaOH treatment (~20% compared to CDTA) and the LM5 and LM6 epitopes were extracted in highest amounts with the strong base (data not shown) [15].

The discovery of β -(1,3)(1,4)-glucan ('mixed linkage glucan' or MLG) epitopes (as recognized by the anti-MLG monoclonal antibody BS-400-3) was also surprising. MLG has been observed outside the Poales in *Equisetum arvense* but, this is to the author's knowledge, the first report of mixed linkage glucan in the lycophytes [15]. In order to confirm the presence of MLG, arrays of *S. moellendorffii* cell wall material were treated prior to BS-400-3 labelling with lichenase, an enzyme that specifically cleaves MLG. This treatment resulted in the complete abolition of BS-400-3 binding and this strongly suggest that the cell walls of *S. moellendorffii* do indeed contain this polysaccharide. (Figure 2). Similarly were the occurrence of homogalacturonan and mannan confirmed by enzymatic removal of the epitope with polymer specific enzymes (Figure 2).

Immunolocalisation of cell wall epitopes. A subset of *S. moellendorffii* cell wall epitopes were immunolocalised in stem sections (Figure 3). The HG epitope recognized by mAb JIM5 was more abundant in the vascular tissues than other stem tissues (Figure 3 A). The binding of the β -(1,4)-galactan mAb LM5 was restricted mostly to the walls of certain phloem cells but in the cortex the LM5 epitope was mainly present in the cell junctions (Figure 3 C, D and E). The α -(1,5)-arabinan epitope recognized by mAb LM6 was widely distributed in cortical cell walls (Figure 3 F and G). In contrast to the pectic epitopes described above, the β -(1,4)-xylan epitope recognized by mAb (LM10) was very widely distributed and labeling was strong in all tissues except the phloem (Figure 3H). Counterstaining with Calcofluor showed that β -glucans were present throughout cell walls and this is consistent with the presence of cellulose (Figure 3 E and G).

Cell wall composition analysis. Sugar composition analysis of the aerial parts of *S. moellendorffii* (Figure 4) revealed that *S. moellendorffii* cell walls are relatively high in mannose and low in galacturonic acid compared to non-commelinid primary cell walls. The abundance of mannose is consistent with a general trend that the cell walls of basal plant species tend to be mannan-rich, and although the prevalence of this polysaccharide suggests an important role in cell wall physiology our present understanding of this is limited. To investigate the properties of the cell wall further, it was analyzed via sequential extraction using CDTA and 4M NaOH and the sugar composition of the resulting extracts was analysed (Figure 4). In order to explore the recalcitrance to extraction observed in the CoMPP analysis the extraction time was extended to 24 h in CDTA and NaOH extractions. Extraction using CDTA resulted in low amount of sugars extracted. The CDTA contained mainly arabinose, galactose, glucose and galacturonic acid, corresponding to the AGP and HG epitopes found in the CoMPP analysis. The NaOH extraction contained large amounts of xylose and mannose, and in both cases were the majority of xylose and mannose using 4M NaOH. Besides xylose and mannose other sugars were observed, but in low quantities, including the pectinous sugars rhamnose, arabinose, galactose and galacturonic acid and glucose.

Together biochemical and immunochemical data suggest that *S. moellendorffii* contains many of the cell wall polymers that are typical of higher plants including MLG. The cell walls of *S. moellendorffii*

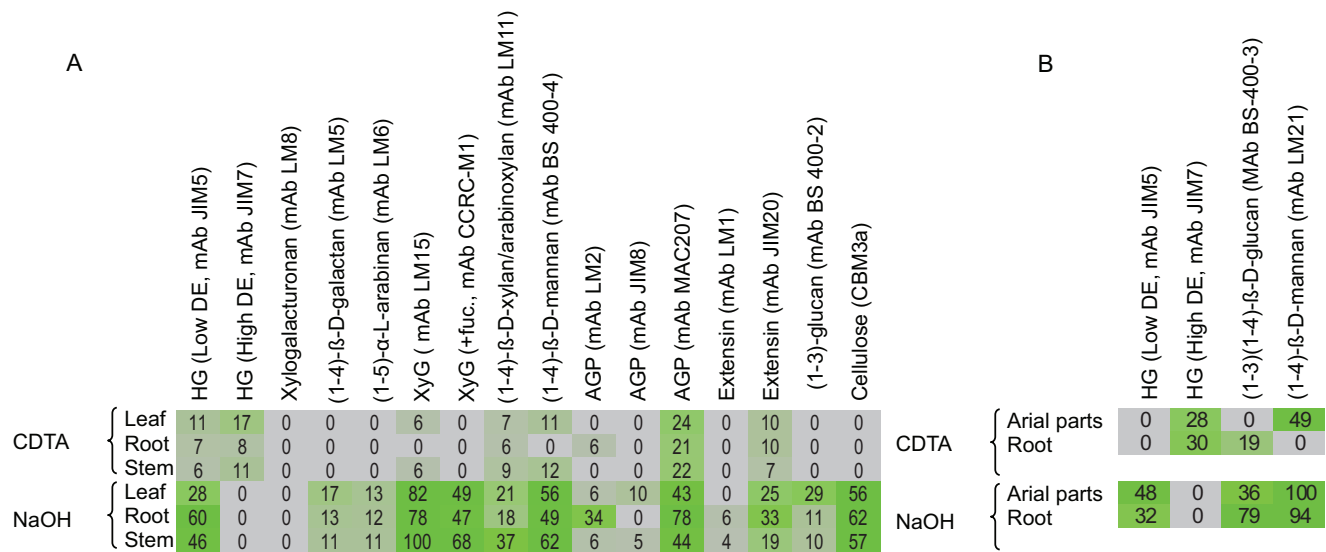


Figure 1. Analysis of polysaccharides in the cell wall of *S. moellendorffii* using the microarray and antibody based CoMPP technique (Moller et al., 2007). Leaves, root and stems (A) and aerial part and roots (B). The heatmap, in which mean spot signals are correlated with colour intensity, shows the relative abundance of cell wall components as extracted using 1,2-diaminocyclohexanetetraacetic acid (CDTA) and NaOH. A low end cut-off value of 5 was used. The same amount of cell wall material (alcohol insoluble residue) was used for each sample. HG, homogalacturonan. XG, xyloglucan. AGP, arabinogalactan protein. This analysis indicates that *S. moellendorffii* has many cell wall polysaccharides in common with higher plants. It is of particular note that in contrast to most other higher plants, HG from *S. moellendorffii* was extracted more readily with NaOH than with CDTA and the occurrence of mixed linkage glucan. However, in common with higher plants but in contrast to *P. patens*, *S. moellendorffii* contained fucosylated XyG.

doi:10.1371/journal.pone.0035846.g001

are distinct from those of seed plants in the abundance of some polymers, the recalcitrance to extraction, as well as their context both at the cellular and cell wall levels. Both ancestral and derived traits are found intermixed in the *S. moellendorffii* cell wall.

The GT repertoire of *S. moellendorffii*. The global survey of GTs in *S. moellendorffii* was based on a BLAST database built

from all GTs in the CAZy database. The entire *S. moellendorffii* and *P. patens* genomes were run against this database. Due to the quality control measures used (see Materials and Methods) the screen for GT-candidates adopted in this analysis is rather conservative and probably lead to very few false positives, but may have underestimated the true number of GTs in *S.*

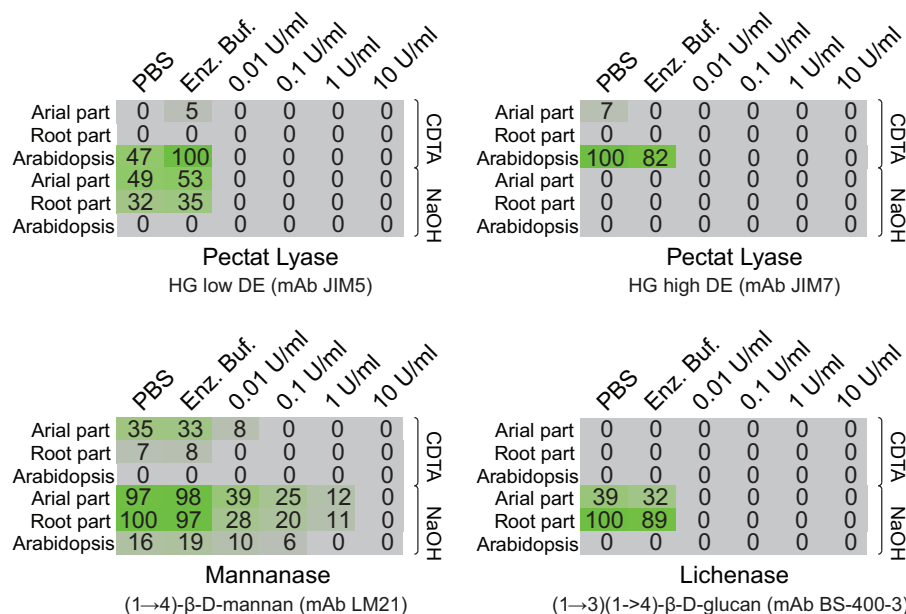


Figure 2. Confirmation that the epitope recognized by homogalacturonan, mixed linkage glucan and mannan specific antibodies are susceptible to enzymatic degradation by enzymes specific for the given polymer. CoMPP arrays were treated with PBS, buffer or increasing amount of enzyme to investigate if the epitope in question could be abolished. In all four cases were the epitope recognized by the antibody susceptible to enzymatic degradation, confirming that the epitope is found in the proposed polymer.

doi:10.1371/journal.pone.0035846.g002

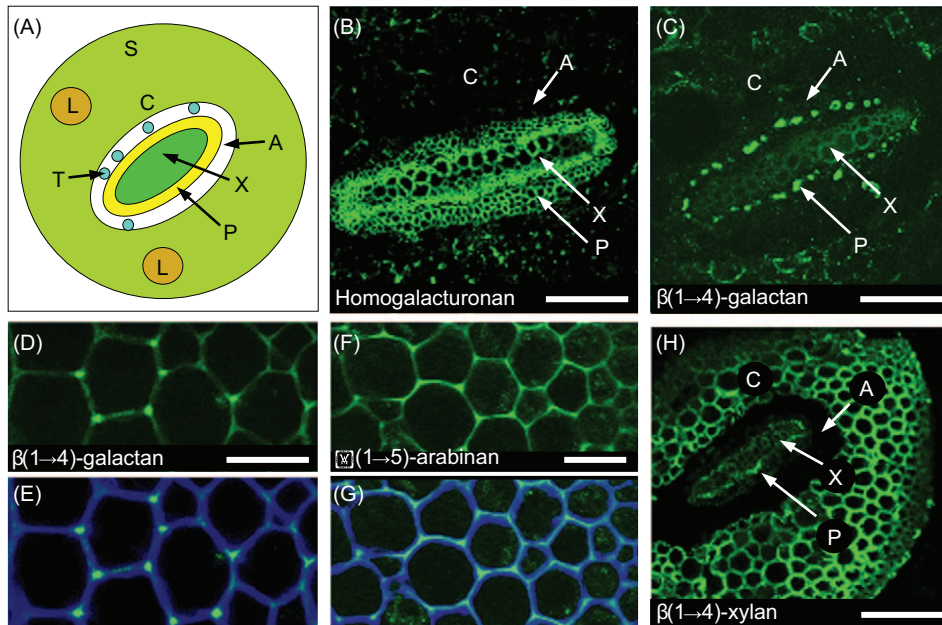


Figure 3. Immunocytochemistry of *S. moellendorffii* stem (A), schematic representation of a section through a *S. moellendorffii* stem showing the location of the cortex (C), sclerenchyma (S), xylem (X), phloem (P), leaf traces (L) and specialized endodermal cells known as trabeculae (T) in the air space (A) separating the stele from the cortex. Indirect immunofluorescence microscopy reveals the location of homogalacturonan (HG, mAb JIM5) (B), β -(1,4)-galactan (mAb LM5) (C), (D) and (E), α -(1,5)-arabinan (mAb LM6) (F) and (G) and β -(1,4)-xylan (mAb LM10) (H). The images in (E) and (G) show counterstaining of the sections shown in (D) and (F) with Calcofluor White that reveals all β -glucan-containing cell walls. In contrast to most higher plants, β -(1,4)-galactan had a highly restricted location in *S. moellendorffii* stems and was located mainly in certain phloem cells (C) and only at the cell junctions of cortical cells (D) and (E). α -(1,5)-arabinan was widely distributed in cortical cell walls, but labelling was restricted to the middle portion of walls rather than throughout their whole width (F) and (G). HG was typically not abundant in the vascular tissues of higher plants but in *S. moellendorffii* HG was located predominantly in the walls of xylem and phloem cells and only present at low levels in other stem tissues. In contrast, labeling of β -(1,4)-xylan was strong in most cell walls apart from those in phloem tissue. Scale bars in (B) and (C) = 140 μ m, in (D) and (F) = 20 μ m, and in (H) = 200 μ m.
doi:10.1371/journal.pone.0035846.g003

moellendorffii. Incorrect gene annotations persist in databases long after the mistakes have been discovered and lead to perpetuation of the errors in annotation of other species. False positives are therefore highly undesired while underestimates (false negatives) are less problematic unless conclusions are made on the total number of GTs in a given species.

The *S. moellendorffii* and *P. patens* GTs were classified according to the CAZy nomenclature and named as follows: GT<family number><clade letter><number>. These names are preceded with a species identifier where required, i.e. Sm for *S. moellendorffii*. Other naming conventions have already been implemented for some families. These are respected and *S. moellendorffii* genes are named after their orthologs in these cases. *S. moellendorffii* and *P. patens* protein IDs and gene names are listed by family and clade in Table S1.

The histogram in Figure 5 comprises 481, 592, 323 and 287 GTs from *Arabidopsis*, rice, *P. patens* and *S. moellendorffii*, respectively. *P. patens* has a rather full set of higher plant polysaccharides [15], but a smaller set of GTs might have been expected since the moss lacks several cell types with specialized walls known from vascular plants. However, the *P. patens* genome is larger with 61% more ORFs than *S. moellendorffii*, so the relatively large number reflects that the *P. patens* set of GTs is generally more redundant.

S. moellendorffii exhibits both ancestral and derived traits in the identified CAZyome. For example, the genomes of *S. moellendorffii* and *P. patens*, but not seed plants, have members of family GT51, which encode murein synthases involved in cell wall biosynthesis

in bacteria, including cyanobacteria from which the *Viridiplantae* chloroplasts arose via an endosymbiotic event (Figure 5). Some of these GTs are penicillin sensitive and interestingly, chloroplast division is sensitive to penicillin in liverworts [16], *P. patens* [17], and *Selaginella nipponica* [18] but not in higher plants. Another striking example is in GT78, containing only two entries namely a mannosylglycerate synthase from *Rhodotermus marinus* (a bacterium) and a homologous partial sequence from *Griffithsia japonicum* (a marine red alga) (Figure 5). Mannosylglycerate is a compatible solute and osmoprotectant [19] produced by many bacteria using either GT78 or, more frequently, GT55 enzymes. GT55 members have not been found within Archaeplastida, but GT78 members with very high similarity to the bacterial mannosylglycerate synthase are present in *S. moellendorffii*, *P. patens* (Table S1) and in the Charophycean green alga *Coleochaete orbicularis* (data not shown).

The majority of the enzymes required for building the vascular plant cell wall are still unknown and hence it is not possible to focus this study on the precise set of CAZy-families required for cell wall synthesis. A global approach has been taken, but families known to be involved in starch biosynthesis, for example, are not included in our analysis with its focus on cell wall biosynthesis. Evidence for involvement in cell wall synthesis in the remaining families is often quite circumstantial. For example, in GT64 loss-of-function of one *Arabidopsis* member (EPC1, At3g55830) has been shown to perturb the cell wall [20,21], but the evidence for a direct involvement of EPC1 in cell wall biosynthesis is not clear.

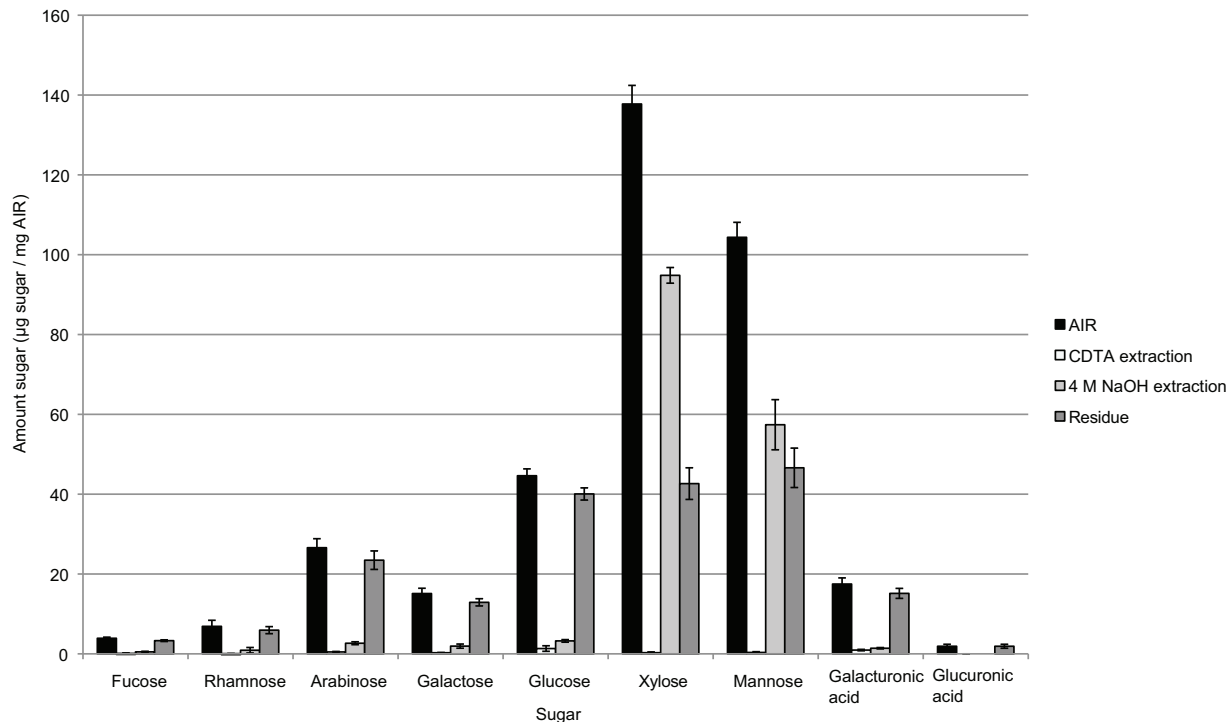


Figure 4. Sequential fractionation of aerial parts of *S. moellendorffii* analyzed with quantitative sugar composition analysis. The recalcitrance to extraction of homogalacturonan as observed in the CoMMP analysis (Figure 1) was confirmed. A general recalcitrance to extraction was observed and only in the cases of xylose and mannose could a majority of the sugars be extracted.
doi:10.1371/journal.pone.0035846.g004

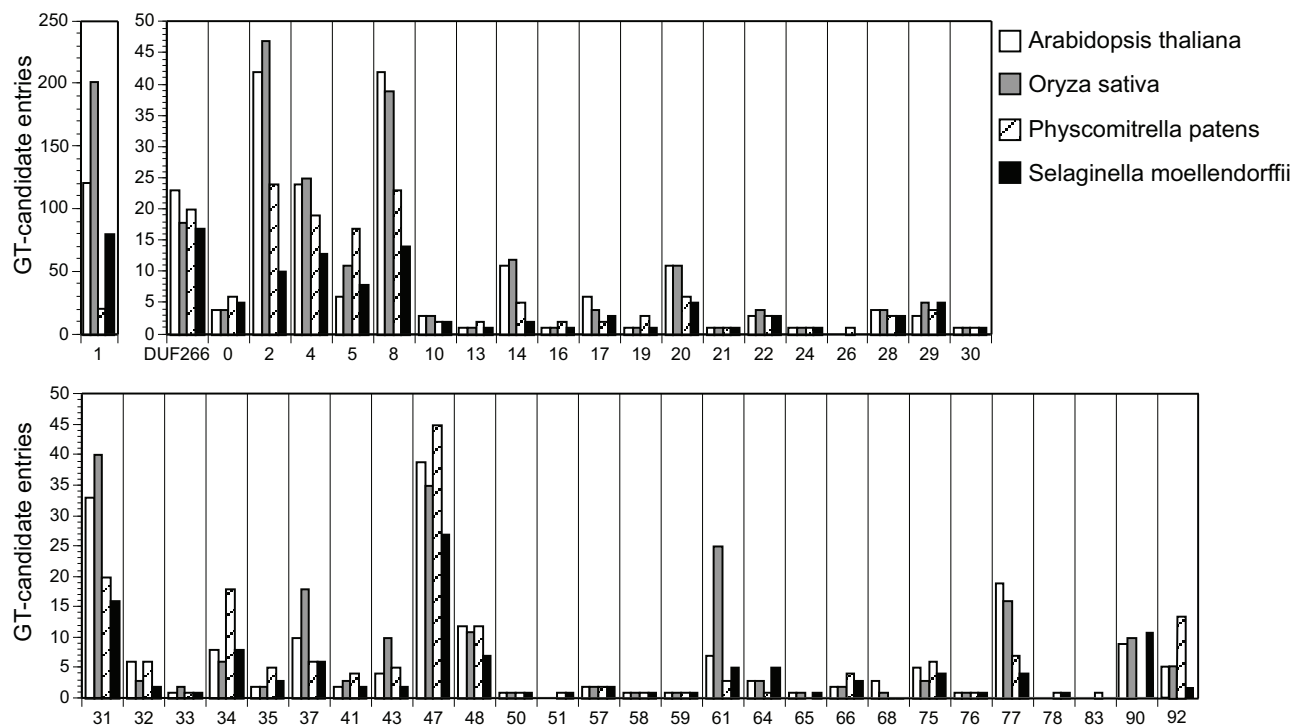


Figure 5. Histogram showing the occurrence of GT-entries across families in the CAZy database and in DUF266-containing genes. *A. thaliana*, white bars; *O. sativa*, grey bars; *P. patens*, hatched bars and *S. moellendorffii*, black bars. GT-numbers on the abscissa are shown in bold for families known or expected to be of particular cell wall relevance. Table S1 lists the JGI protein IDs for both *P. patens* and *S. moellendorffii* and gene names for the latter. Sequences are accessible via the JGI websites and at NCBI.
doi:10.1371/journal.pone.0035846.g005

GT64 is divided into three groups with *S. moellendorffii* members in all three (Figure S1).

Glycosyltransferases involved in cell wall biosynthesis

The cellulose synthase superfamily. The CAZy GT2 family is very large, including many enzymes not found in plants such as chitin synthase, hyaluronan synthase and others involved in protein N-glycosylation such as dolichyl phosphate β -glycosyl transferase. We have limited our analysis to the *CELLULOSE SYNTHASE* (*CESA*) superfamily, which includes proteins known and proposed to be involved in plant cell wall biosynthesis. The *CES* superfamily took its name from the *CESA* genes, which encode the catalytic subunits of cellulose synthase, and includes nine families of *CELLULOSE SYNTHASE-LIKE* (*CSL*) genes proposed to encode the GTs that synthesize the backbones of other cell wall polysaccharides [22]. The *CESAs* and *CSLs* are integral membrane proteins that are active in the plasma membrane (*CESAs*) or Golgi (at least some *CSLs*). Members of the *S. moellendorffii* *CESA* superfamily have been named using the established convention [22]. For historical reasons, we have included the callose synthases from the GT48 family in this section. Based on the ubiquitous presence of callose in *in vitro* cellulose biosynthesis assays, it was once hypothesized that cellulose and callose were synthesized by the same enzyme [23,24]. However, it was subsequently discovered that callose synthases are similar to the yeast β -1,3-glucan synthase Fks [25], which is a member of GT48. All members of family GT48 are expected to be β -1,3-glucan synthases so no clade names are proposed for GT48 (Figure S2).

Analysis of the *CESA* gene superfamily in *S. moellendorffii* revealed many similarities with that of *P. patens*, including 1) absence of orthologs of the spermatophyte *CESA* genes that are specialized for primary and secondary cell wall biosynthesis, 2) presence of genes with similarity to cyanobacterial and red algal *CESA* genes, which are absent in spermatophytes, and 3) presence of only three *CSL* gene families, compared to seven or eight in seed plants (Figures 4, S3 and S4). The *CESA*, *CSLA*, *CSLC* and *CSLD* gene families of *S. moellendorffii* are smaller than those of either *P. patens* or seed plants and phylogenetic analysis revealed that each of these families diversified independently within the bryopsid, lycopod and seed plant lineages.

In Arabidopsis, the proteins encoded by three *CESA* genes (AtCESA4, 7, 8) are required for secondary cell wall biosynthesis in vascular tissue [26]. Proteins encoded by three different *CESA* genes (AtCESA1, 3 and either AtCESA2, 5, 6 or 9) are required for primary cell wall biosynthesis [27,28]. All spermatophytes examined to date have orthologs of at least three primary wall and three secondary wall Arabidopsis *CESA* genes [29,30]. *P. patens* lacks orthologs of these specialized *CESA* genes, as might be expected for a non-vascular plant [31]. Surprisingly, these orthologs are also lacking in *S. moellendorffii* (Figure S2) indicating that *CESA* diversification and specialization for primary and secondary cell wall biosynthesis was not necessary for the evolution of vascular tissue. This observation and the observation that monolignol biosynthesis, at least partly, evolved convergently with euphyllophytes monolignol biosynthesis lend genomic support to structural observations suggesting that the vascular tissue in *S. moellendorffii* is not homologous to euphyllophytes vascular tissue [32,33]. Analysis of *CESA* genes from basal euphyllophytes will be required to better understand the relationship between tracheary element/secondary cell wall evolution and *CESA* diversification.

The *CESA* proteins of cyanobacteria [34] and the red alga *Porphyra yezoensis* [35] lack three defined regions (i.e. Zn-binding domain, Plant Conserved Region, and Class Specific Region) that

are conserved in streptophyte *CESAs*. Only the streptophyte-type *CESAs* are known to form rosette terminal complexes [36] and the *P. yezoensis* *CESAs* form linear terminal complexes [37]. In addition to seven streptophyte-type *CESA* sequences, *P. patens* have one gene that is similar to cyanobacterial and red algal *CESAs* and sequences from ascomycete fungi. Similar sequences are not found in spermatophytes, but occur in the genome sequence of *S. moellendorffii* where they exist as a tandem duplicate and a group of four tandem genes. Some of the genes in the tandem repeats are pseudogenes and one, GT2A5/6, appears to be a fusion between two genes. So five full length GT2A's can be identified in the genome of *S. moellendorffii* (Figure 6). Although similar to *CESAs* from early-divergent taxa, these GT2 genes have undergone substantial diversification in *S. moellendorffii*, perhaps reflecting acquisition of a novel function.

The *CESA* gene superfamily also includes *CSLA* genes, which encode (gluco)mannan synthases [38,14]; *CSLC* genes, which putatively encode XyG synthases [39]; *CSLF* [40] and *CSLH* [41] genes, which encode mixed-linkage β -glucan synthases; and five additional families of *CSL* genes, which have been proposed to encode the synthases that polymerize other cell wall polysaccharides [22,42]. Only the *CSLA*, *CSLC* and *CSLD* gene families are represented in the genomes of *P. patens* [31] and *S. moellendorffii* (Table S1). This indicates that the *CSLB*, *CSLE*, *CSLF*, *CSLG*, *CSLH* and *CSLJ* gene families, and presumably novel cell wall polysaccharides synthesized by their encoded proteins, originated after divergence of the lycopsids from the vascular plant lineage. This observations is quite enigmatic as the biochemical analysis of *S. moellendorffii* cell walls revealed the presence of MLG. In the Poales, MLG is a product of the GTs encoded by *CslFs* and *CslHs*, but orthologues of these genes are not found in *S. moellendorffii*. This implies that the MLG in *S. moellendorffii* is produced by a different biosynthetic route than in Poales species and hence has emerged at least twice in embryophytes by convergent evolution. This raises the question of which *S. moellendorffii* GTs are responsible for the production of MLG. No obvious candidates emerge from the GTs identified in this study, indicating they may not be included in the CAZy database. Phylogenetic analysis revealed independent diversification of the *CSLC* and *CSLD* (Figure S4 and S5) gene families within the moss and lycophte lineages. Although a catalog of *CESA* superfamily genes from sequenced plant genomes reported larger numbers of *CESA* and *CSL* genes in *S. moellendorffii* [43], these include allelic variants and two pseudogenes.

Is *CSL* function conserved among the major land plant lineages? Mannan/glucomannan synthase activity has been demonstrated directly for *P. patens* *CSLA* proteins [14], indicating that *CSLA* function is conserved. In contrast, the role of *CSLC* proteins in the synthesis of the XyG backbone has been examined only in seed plants and only shown for one member of *CSLC* [39]. As shown in this and previous studies, both mannan and XyG polymers are present in *P. patens*. *CSLD* proteins are involved in tip growth and early cell differentiation [44,45]. Similarity to *CESAs* and complementation experiments have led to suggestion that *CSLDs* may be glucan synthases [46,47] but heterologous expression of *CSLDs* in tobacco indicated that they have mannan synthase activity [44].

Xyloglucan and galactomannan. The biosynthesis of XyG is well understood with members of the GT2 *CSLC*-clade, GT34, GT37 and GT47 families implicated. The Arabidopsis XyG xylosyltransferase activities (At3g62720 and At4g02500, [48]) are placed in clade A of family GT34 with one rice and the *S. moellendorffii* sequences GT34A1 and GT34A2 (Figure S6). Surprisingly, no *P. patens* sequences could be found in clade A. The ancestral nodes of GT34 are not resolved but clades, named

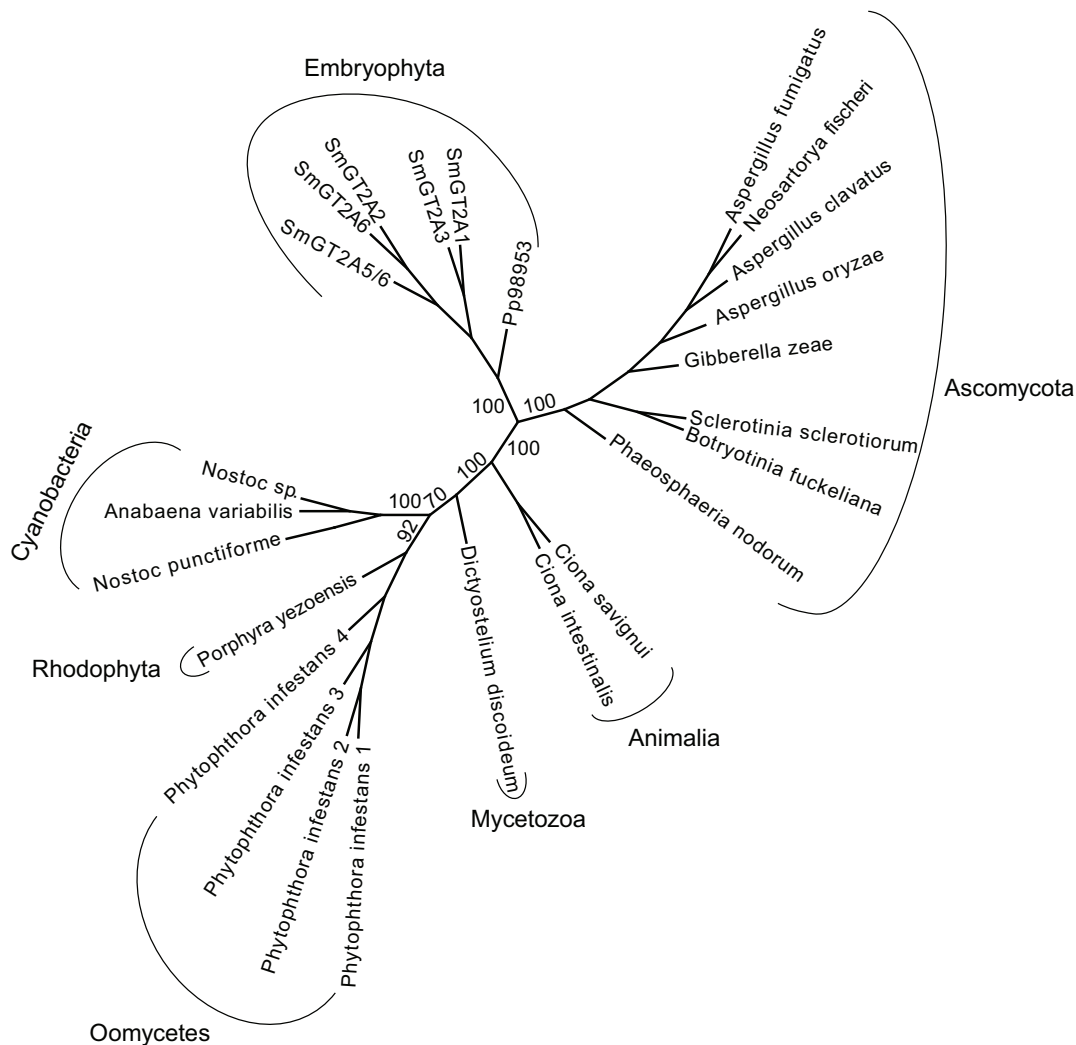


Figure 6. Phylogenetic tree of GT2. The genomes of *P. patens* and *S. moellendorffii* include GT2 sequences with no orthologs among seed plants. Putative cellulose synthases from cyanobacteria and the red alga *Porphyra yezoensis* are the most similar sequences from within the green plant lineage. Other similar sequences include known cellulose synthases from oomycetes, the slime mold *Dictyostelium discoideum*, and two species of tunicates, e.g. sea squirts (animalia), and ascomycete sequences of unknown function.

A through to D, are distinguishable in the tree and supported by high bootstrap values. The *Arabidopsis* proteins in clade C may also be XyG xylosyltransferases [49], and that could suggest that the *P. patens* proteins in clade D are xyloglycan xylosyltransferase as well. A much more detailed study of the proteins in clades A, C, and D will be required to clarify the divergence of these enzymes, which may all catalyze the same reaction. *P. patens* entries are found in clade B, which includes galactomannan galactosyltransferases. The antibody BS-400-4 and LM21 used to detect mannans in *P. patens* and *S. moellendorffii*, respectively, has a rather broad specificity, hence binds to both β -(1,4)-D-mannan and galacto- β -(1,4) -D-mannan [50,51] (Petersen HL and Willats WG, unpublished). The galactomannan structure has thus not been resolved in *P. patens* or *S. moellendorffii*, so the sequence analysis cannot be correlated with biochemical evidence.

Three genes, At2g20370 (MUR3), At2g32740 and At5g62220, proposed to be involved in galactosylation of XyG, are placed in clade A in GT47 [52,53](Figure 7). Mur3 and At2g32740 are placed in a low-resolution subclade along with *S. moellendorffii*, *P. patens* and rice sequences (Figure 7). At5g62220 is, along with only

S. moellendorffii (SmGT47A10 and SmGT47A4) and rice, placed in a resolved subtree (Figure 7). Which galactose in XyG that At5g62220 transfers is unknown. As there are no *P. patens* orthologs of At5g62220 it could be proposed that it transfers galactose to the ultimate α -1,6-xylose in the tetra or pentamer subunit in XyG as this galactose is substituted with galacturonic acid or arabinose in *P. patens* [12].

The occurrence of fucose in XyG in different seed plant families has a bearing on the interpretation of the phylogenetic relationships of family GT37 for *S. moellendorffii* (Figure 8). GT37 comprises several Arabidopsis fucosyltransferases, of which one and only one, AtFUT1, is a XyG fucosyltransferase [54]. Two other members of the GT37 FUT clade have been implicated in fucosylation of AGP [55]. Both, failure to detect terminal fucose by linkage analysis in rice and oat cell walls [56,57,58] and orthology of GT37-member Os02g0764400 with AtFUT4 and 5, which do not fucosylate XyG [59], supports the hypothesis that grass XyG is not fucosylated [60]. However, fucosylated XyG has been detected by radiolabeling in the grass *Festuca arundinacea* [61], so fucosylation persists at least in some genera of the grass family or possibly occurs

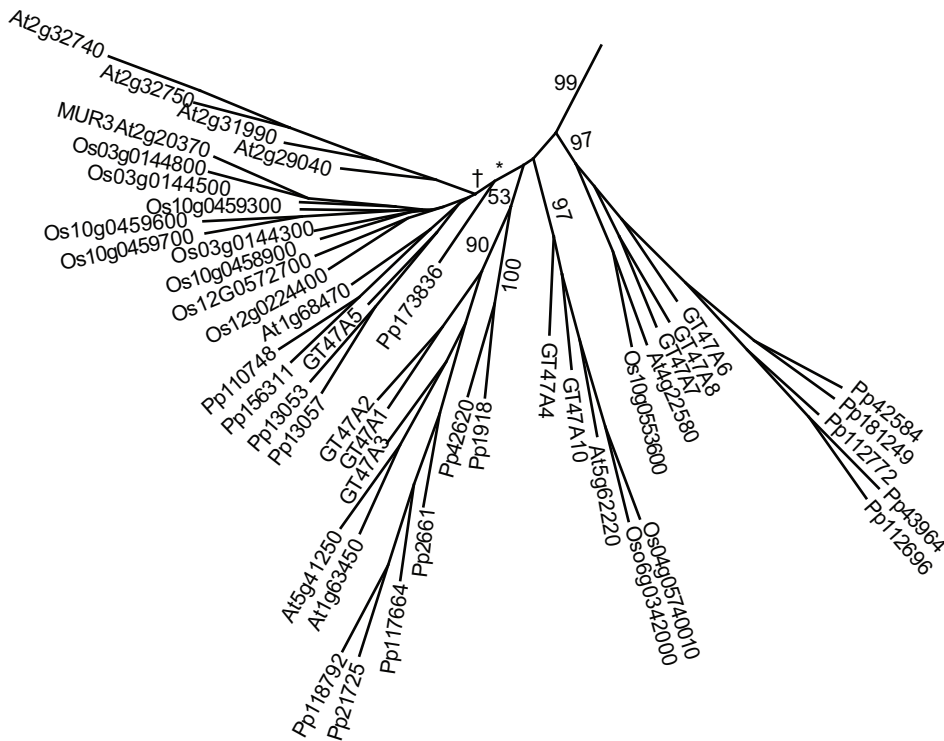


Figure 7. Phylogenetic tree of the A-clade of GT47, which includes the known XyG galactosyltransferase MUR3 and putative galactosyltransferases GT13 and GT18. The A-clade includes a mixed subclade with an overrepresentation of *P. patens* and *S. moellendorffii* sequences, a subclade containing At5g62220 with no *P. patens* members, and a subclade containing At5g41250 with no rice members, but is otherwise poorly resolved. Naming according to Li et al (2004): At5g62220 (GT18), At2g32740 (GT13) and At1g68470 (GT17). GT47 clade A members that are not included due to incomplete sequences include At4g13990, Pp156314 and Pp201625. Branching distal to † is not resolved.

transiently during biosynthesis as recently reported for xylosylation of XyG [62]. XyG of the moss *P. patens* is of the non-fucosylated type as judged by probing with monoclonal antibody CCRCM1, which recognize fucosylated XyG [11]. This antibody binds to *S. moellendorffii* cell wall extracts (Results not shown), and its GT37 members could include fucosyltransferases, but *S. moellendorffii* only contains GT37A members, a clade with only *P. patens* and *S. moellendorffii* GTs, and thus no ortholog of AtFUT1.

Xylan

Xylans are the other major class of hemicellulosic polysaccharides and some progress has recently been made towards understanding its biosynthesis. The xylan backbones of several different species of gymnosperms and angiosperms have been shown to have a complex reducing end oligosaccharide hypothesized to be a primer or terminator for the synthesis [63]. Synthesis of the oligosaccharide in *Arabidopsis* is believed to involve IRX8 and PARVUS (GT8) and IRX7 (Figure S7) all of which appear to have orthologs in *S. moellendorffii* suggesting that it also makes the reducing end oligosaccharide. *S. moellendorffii* has only one member of the GATL group of GT8, to which PARVUS belongs (Figure S8). In contrast, *Arabidopsis* has 10 GATL members. Since the *parvus* mutant has a very strong phenotype, these 10 GATL members do not seem to be redundant [64]. The *S. moellendorffii* GATL is not a clear ortholog of PARVUS so we cannot safely conclude that it has the same function as PARVUS in xylan biosynthesis. The GTs responsible for elongation of the xylan backbone in *Arabidopsis* are likely to be IRX9/IRX9L and IRX14/IRX14L, which are located in different subgroups of GT43 (Figure S9)[65]. The close homologs IRX10 and IRX10L

of GT47 are also involved in xylan backbone biosynthesis [66,67]. Both *S. moellendorffii* and *P. patens* have IRX14/IRX14L and IRX10 orthologs. *S. moellendorffii* lacks direct orthologs of IRX9 but has orthologs of IRX9L (At1g27600). Both IRX9 and IRX14 are highly expressed in tissues containing secondary cell wall, but IRX14 follows this pattern less stringently [65], and the IRX9L gene in Arabidopsis is ubiquitous expressed according to *in silico* expression profiling [68]. Thus, it can be hypothesized that the seed plants have evolved a special isoform of IRX9, which is specific for secondary cell wall biosynthesis, while *S. moellendorffii* and *P. patens* only have the original isoform without a tissue specific role. This lack of GT specialization in *S. moellendorffii* is in line with the lack of specialized forms of CESA for secondary walls in this species (see discussion on cellulose synthase superfamily). The GTs, glucuronosyltransferases (GUX), that add glucuronic and methyl-glucuronic acid to xylan have been identified, [69,70] as member of GT8 within a clade containing starch initiation proteins [71,64]. Both *S. moellendorffii* and *P. Patens* have GUX orthologs but they are placed basal in the phylogenetic tree so possible specialization of GUX in higher plants has apparently not occurred in *P. patens* and *S. moellendorffii* (Figure S10). An alternative possibility is that all the higher plant GUX proteins have the same activity. GT61 is a family with many GTs in rice, several of which are coexpressed with genes predicted using *in silico* analysis to be involved in arabinoxylan biosynthesis [72,70]. All four species have members in the A and B-clades of GT61 (Figure S11 and not shown). No activities are known for the A-clade whereas the distantly related B-clade contains the Arabidopsis β -1,2-xylosyltransferase involved in protein N-glycosylation and an apparent ortholog from each species [73]. The C and D

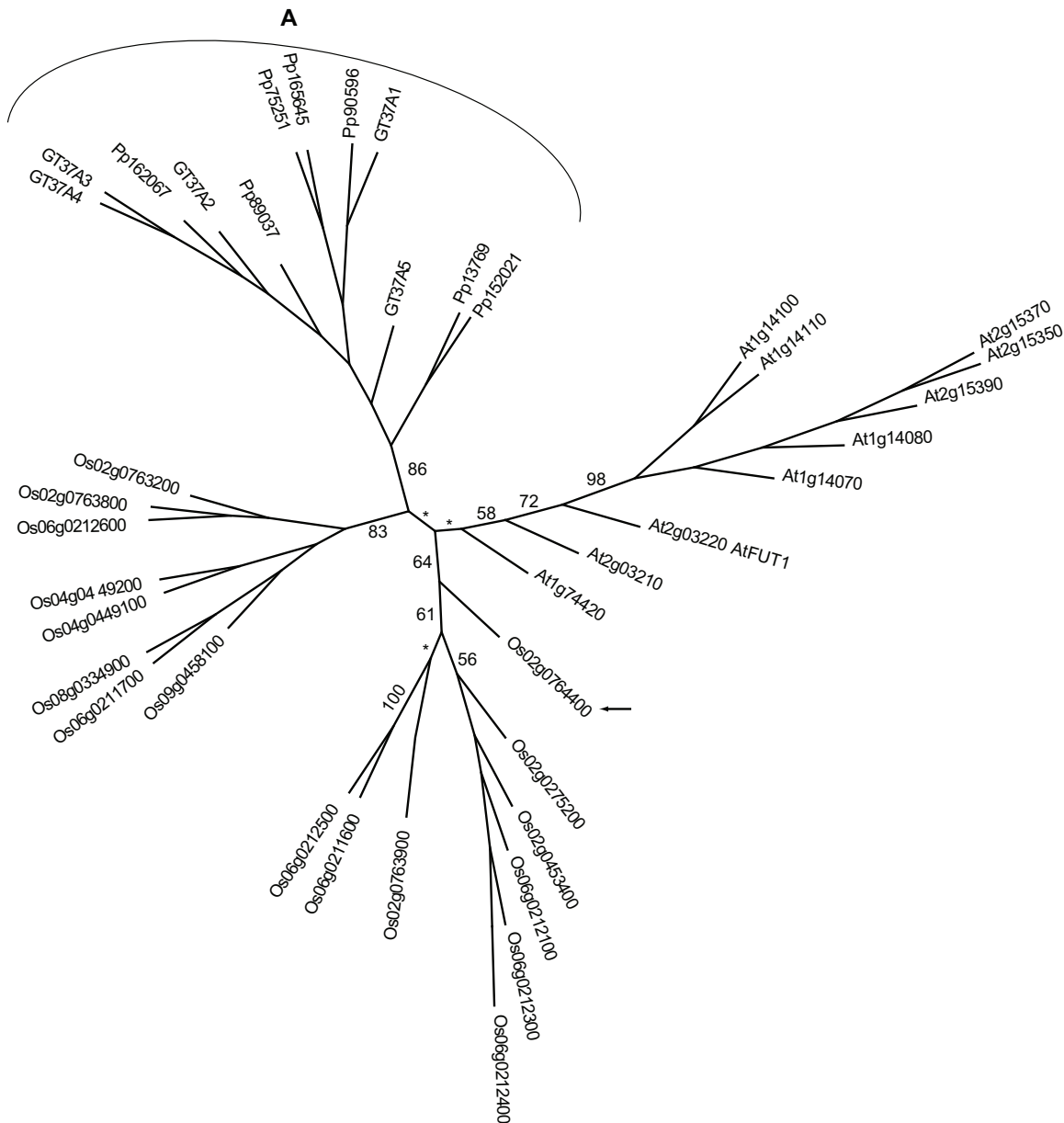


Figure 8. Phylogenetic tree of GT37. The lower plant clade A is clearly resolved while the structure of the remainder of the tree is ambiguous. Known activity is xyloglucan α -1,2-fucosyltransferase AtFUT1. The other members of the Arabidopsis group are known not to fucosylate xyloglucan, but are likely to be fucosyltransferases (and are named AtFUT2-10). Os02g0764400 (arrow) is the most similar rice sequence to AtFUT1 although orthology is not clearly suggested by the clade structure. Adding *Brachypodium distachyon* GT37 sequences to the analysis (data not shown) will move Os02g0764400 and its *Brachypodium* ortholog into the putative Arabidopsis clade clearly showing that the clade structure is not stable. GT37 candidates that are not included due to incomplete sequences include At2g15350, Pp23092, Pp45180, Pp45146, Pp89038 and SmGT37A6.
doi:10.1371/journal.pone.0035846.g008

clade of family GT61 is greatly expanded in grasses, with 18 rice entries compared to two for *Arabidopsis* and no members from either *P. patens* nor *S. moellendorffii*, suggesting that these enzymes catalyze a reaction that predominates in grasses. Mitchell et al. first proposed that the proteins could be responsible for β -1-2 xylose transfer onto arabinose sidechains of xylan [74,72]. Recently an activity has been identified in GT61, a wheat clade C ortholog was shown to be a xylan arabinosyltransferase [75]. The LM10 antibody, which shows that xylan is present in *S. moellendorffii*, binds to unsubstituted or lowly substituted xylan. Determination of the fine structure of *S. moellendorffii* xylan would be helpful to the deductions of activities in GT61.

β -1,4-xylan is not known to be present in algae and based on the genetic complexity with the involvement of at least seven different activities for xylan biosynthesis, xylan biosynthesis is illustrative for the evolutionary jump that the earliest plants made in order to synthesize a cell wall suited for the terrestrial environment.

Pectin

Pectin is the most complex cell wall polysaccharides, and their biosynthetic pathways are poorly understood [76,77,78,79,80]. The synthesis of the α -1,4-linked galacturonan backbone of homogalacturonan is catalyzed by GAUT members of family GT8, the galacturonosyl transferases. Presumably, the galactur-

onan backbones of RG-II and xylogalacturonan are synthesized by members of the GAUT family as well. Genes similar to the GAUT1 homogalacturonan synthase gene (At3g61130) are present in *S. moellendorffii* and *P. patens*, consistent with the presence of homogalacturonan in both species [81,82]. The GAUT group within GT8 has subgroups that are missing in *S. moellendorffii* and *P. patens* or are missing only in *P. patens* (Figure S8). Of special interest is the clade containing AtGAUT7 (At2g38650), which has only angiosperm members. AtGAUT7 has been proposed to be part of a complex of GTs involved in homogalacturonan biosynthesis, even though homogalacturonan synthase activity has not been shown for AtGAUT7 [76]. Quasimodo1 (At3g25140) has been shown in Arabidopsis to be involved in homogalacturonan synthesis by mutation analysis [83]. The exact activity of Quasimodo1 has not been established and pleiotropic effects have been observed in the mutant [84]. *P. patens* does not have an ortholog of Quasimodo1 whereas *S. moellendorffii* does, pointing to a possible gradual evolution of homogalacturonan biosynthesis with complex formation requiring GAUT7 in angiosperms as the most recent development.

Family GT47 comprises, inter alia, GTs involved in pectin side-chain synthesis [85,86]. A homolog of putative arabinan arabinosyltransferase ARAD1 was found in *S. moellendorffii* but not *P. patens* (Figure S7). As arabinan is found in *P. patens*, other members of the GT47B clade could be redundant to ARAD1, supported by the presence of residual arabinan in *arad1*. The arabidopsis xylosyltransferase XGD1 involved in xylogalacturonan biosynthesis lacks orthologs in both *P. patens* and *S. moellendorffii*. Interestingly, the most similar *S. moellendorffii* gene, GT47C1, occupies a separate subclade, which also has a *P. patens* member.

Only one activity involved in RG-II biosynthesis has been identified, the RGXT xylosyltransferases of family GT77 [87,88]. While Arabidopsis has three or four RGXT paralogs [87,88] in the B-clade, rice, *P. patens* and *S. moellendorffii* have a single RGXT ortholog (Figure 9 and Figure S12). Although the complete RG-II molecule has never been detected in a moss, the presence of RGXT orthologs and finding of diagnostic sugars of RG-II yield evidence for presence of at least a precursor oligosaccharide homologous to RG-II [89].

Extensins and AGPs

Extensins are structural wall polymers comprising a protein backbone that is glycosylated on contiguous hydroproline residues [90]. The glycans are typically $(\beta\text{-}1,2\text{-Araf})_3$, where most side chains carry a terminal $\alpha\text{-}1,3\text{-Araf}$. The RRACs in GT77A and Xeg113 of clade C are putative β -arabinosyltransferases transferring the second and third β -arabinose, respectively (Figure 9) [91,92,93]. Extensins are regarded as ancient because a quite similar group of proteins make up the wall of the chlorophyte alga *Chlamydomonas reinhardtii* albeit with a somewhat richer set of glycan structures [94] and both *P. patens* and *S. moellendorffii* have GT77 clade-A and C members.

AGPs are cell wall glycoproteins that are attached to the plasma membrane through glycosylphosphatidylinositol lipid anchors, reviewed recently and suggested clade naming convention is followed here [95,96,97]. It is generally assumed that the plant members of family GT31 comprise activities involved in adding galactosyl units to O-glycans of AGPs. This is based on the observations that animal GT31 all appear to transfer hexosyl monosaccharides in $\beta\text{-}1,3$ [98] and preliminary studies have shown galactosyltransferase activity for several Arabidopsis members of the B-clade of GT31 (N. Geshi, E. Knoch, and H.V. Scheller, unpublished) and one member of the C-clade (I. Damager, C. Toft Madsen, B. Petersen, J. Harholt and P.

Ulvskov, unpublished). Only a single plant member of GT31, At1g26810 belonging to the B-clade, has been functionally characterized and shown to be a galactosyltransferase involved in the synthesis of the Fuc $\alpha\text{-}1,4$ (Gal $\beta\text{-}1,3$) GlcNAc-R Lea structure of N-glycans [97]. At1g26810 belongs to the B clade of GT31 comprising a pfam 01762 galactosyltransferase domain, and a galactoside binding lectin domain pfam 00337, yet good evidence was presented that the B-clade sequences are not paralogs [97]. At1g26810 alone is implicated in Lea structure synthesis. This raises a problem in gene naming. The *S. moellendorffii* gene, prot ID = 183099, which is closest to At1g26810 is probably closer to At3g06440 known not to be involved in Lea synthesis. *S. moellendorffii* has orthologs to both the ($\alpha\text{-}1,3\text{-FucTs}$) in GT10 and to FucTc ($\alpha\text{-}1,4\text{-FucT}$) which takes part in Lea synthesis. This raises substantial doubt as to the presence of the Lea structure in *S. moellendorffii* and we have named the gene closest to AtGALT1 GT31B1 awaiting biochemical evidence (see Figure S13 and S14).

A large group of mammalian members of GT31 are the fringe GTs, which are N-acetyl-glucosaminyltransferases. The fringe-related GTs in GT31 form clade C, which is distinct from clades B to F and is presented as a separate tree (Figure S15).

Two of the arabidopsis sequences outside CAZy identified as GT-candidates, At5g11730 and At3g21310, carried domain of unknown function DUF266 [99]. LOPIT proteomics experiments have localized At5g11730 to the Golgi [100]. Recently, DUF266s were found in a bioinformatics search for putative GTs in arabidopsis [101], and the brittle culm mutation BC10 in rice was linked to a lesion in a gene with a DUF266-domain that encodes a GT potentially involved in AGP-biosynthesis [102]. BC10 is part of DUF266 Clade A comprising three subclades with unresolved common ancestry (Figure S16). BC10 has orthologs in both *S. moellendorffii* and *P. patens*. No CAZy family exists for the DUF266s yet due to the lack of biochemical evidence. Consequently, we do not propose names for the GTs at this point. The numbers in both the table and the tree are JGI protein IDs.

Concluding remarks

Most of the cell wall polymers found in angiosperms are reported here to be present in both *S. moellendorffii* and *P. patens*. However, there were notable differences observed in the relative abundance of specific polymers and in some aspects of cell wall architecture. In contrast to most angiosperm cell walls, pectin appeared to be most abundant in the vascular tissue whilst xylan was observed mainly in cortical cell walls. Xylan and mannan appears to be the quantitatively most abundant polymers, with minor amounts of pectin. MLG was also found in the cell wall of *S. moellendorffii*, intriguing since orthologues of neither *CslF* or *CslH* were identified in the genome of *S. moellendorffii*. These findings which suggest the evolution of diverse strategies for utilizing specific cell wall components may have some relevance with regard to the remodeling of cell wall structures that is required for optimizing plant biomass for biofuel use.

The comparative genomic approach used in the building of the *S. moellendorffii* CAZyome gave rise to interesting discussions but no clear functionality could be assigned, most likely due to the overlapping presence of cell wall polymers in the four species analyzed. We have presented a gene naming convention for GTs similar to that already in place for the model grass *Brachypodium distachyon*. The next step would be to extend the analysis and include Charophycean and Chlorophycean algae. As an example of this were the cell wall related GT2 in some Chlorophycean green algae analyzed [103]. Both CslA/C orthologs were found, in agreement with the presence of mannans, and no streptophyte-

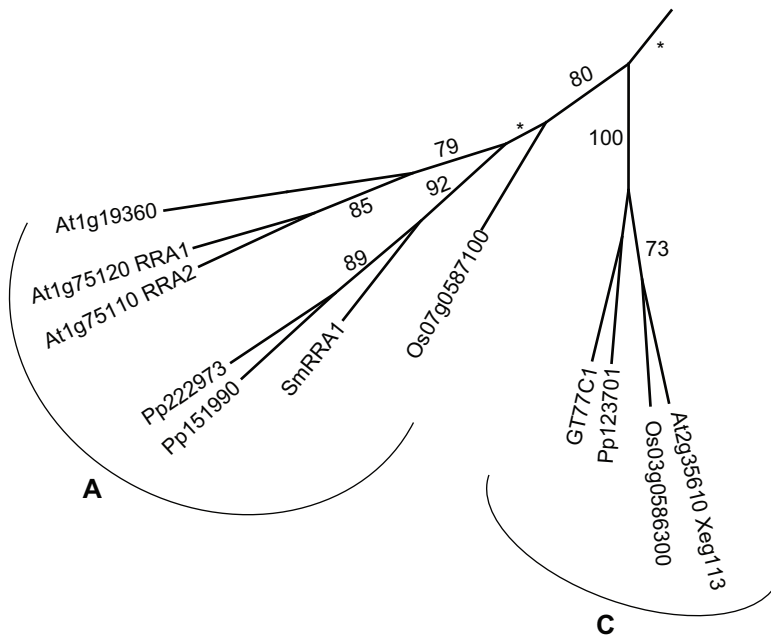


Figure 9. The A and C clades of GT77. The A clade holds the genes affected in the *reduced residual arabinose (rra)* mutants considered to be impaired in extensin arabinosylation. The arabidopsis member of the C-clade has recently also been implied in extensin arabinosylation based on analysis of the *xeg113*-mutant. Both clades have members from all four taxa. The full GT77 tree is Fig. S10.
doi:10.1371/journal.pone.0035846.g009

type CESA was identified. But additional knowledge could be gained on for example extensin biosynthesis if the analysis was broadened to include the whole CAZyome.

Materials and Methods

Plant material

Selaginella moellendorffii for CoMPP and microscopy was obtained from Plant Delights Nursery, Inc., Raleigh, NC, USA and maintained under humid conditions with low light intensity. Alcohol insoluble residue (AIR) was made by freezing tissue in liquid nitrogen and homogenizing using a TissueLyser (Sigma-Aldrich). Five volumes of 70% EtOH was added to the fine powder and incubated with rotation for 1 h. The material was spun down and the EtOH changed before another hour of incubation. This procedure was repeated five times before a final wash with acetone, and the samples were left to air dry.

CoMPP

CoMPP was carried out essentially as previously described [11]. Cell wall polymers were sequentially extracted from 10 mg of AIR with 50 mM diamino-cyclo-hexane-tetra-acetic acid (CDTA), pH 7.5, and 4 M NaOH with 0.1% v/v NaBH₄, and extractions printed in three dilutions and three replicates giving a total of 9 spots per sample. The entire experiment was repeated three times and the average of these three experiments is reported. The data were converted into a heatmap format using the online BAR heatmap tool (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper.cgi). All JIM, LM and Mac antibodies as well as CBM3a were obtained from PlantProbes, UK, and BS 400-2 and -4 was obtained from Biosupplies, Australia. Secondary alkaline phosphatase anti-Rat and anti-Mouse antibodies were obtained from Sigma-Aldrich.

Fluorescence microscopy

Transverse sections of fresh *S. moellendorffii* stems were cut using a vibrating knife microtome (Vibratome, St. Louis, MO 63134, USA). Sections were incubated for 1 h in primary antibody diluted 1:10 in phosphate buffered saline (PBS, 140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, pH 7.2) containing 5% (w/v) fat-free milk powder (MP/PBS). mAbs were used with specificity for HG (JIM5) [104], α -(1,4) xylan (LM10) [105], β -(1,4)-galactan (LM5) [106] and α -(1,5)-arabinan (LM6) [107]. After washing in PBS, sections were incubated for 1 h in fluorescein isothiocyanate (FITC)-conjugated secondary antibody (anti-rat-FITC, Sigma, Poole, UK). After washing in PBS, sections were mounted in antifade agent (CitiFluor, Leicester, UK) and examined using a laser scanning confocal microscope (LSCM510, Carl Zeiss, Switzerland).

Sequential extraction and sugar composition analysis

AIR was sequentially extracted with 50 mM diamino-cyclo-hexane-tetra-acetic acid (CDTA), pH 7.5, and 4 M NaOH with 0.1% v/v NaBH₄. Approximately 30 mg AIR was extracted with 1 ml of solvent. Both extractions were carried out at room temperature for 24 hours with 1400 rpm shaking. AIR, residue and extractions were hydrolysed with trifluoroacetic acid and released sugars were detected and quantified as previously described [108].

Proteomes and database creation

The filtered models proteome of *S. moellendorffii* was acquired from the JGI web-site. Protein sequences of the CAZy-database were downloaded from NCBI and used for generating a CAZy-blast database. The Arabidopsis proteome was downloaded from TAIR and all genes already in CAZy were subtracted and an Arabidopsis blast database depleted of known GTs was prepared.

The screen

The *S. moellendorffii* proteome was blasted against both the CAZy-blast-database and the GT-depleted arabidopsis database using an e-value of 10^{-25} as threshold. Hits to CAZy for which a better (e-value smaller by a factor of five or more) hit was found in the GT-depleted Arabidopsis database were eliminated. The remaining candidates were then screened by an rpsblast against the conserved domain database CDD [109]. Both rpsblast and the CDD are made available by NCBI. CDD does not cover all families, for those families not covered, fold prediction by the Phyre-server was employed [110]. We have attempted not to give names to obvious fragments and pseudogenes, so the final threshold to pass was based on manual screening of alignments of *S. moellendorffii*, Arabidopsis, rice and *P. patens* sequences in the given family. Erroneous *in silico* sequence predictions, mostly wrongly predicted start site and intron/exon borders, were manually edited and corrections were uploaded to JGI. New protein ID number was assigned when editing so please refer to Table S1 for file containing all corrected protein IDs used in present study. Uncorrected sequences can still be obtained by BLAST on JGI's Selaginella site (<http://genome.jgi-psf.org/cgi-bin/runAlignment?db = Selmol&advanced = 1>).

The GT2 and GT8 families were handled individually. GT8 is a very divergent family and the analysis was limited to the GAUT, GATL and GUX groups of genes. Likewise analysis of the very large GT2 family was limited to identification of members of the plant CESA superfamily and genes with similarity to cyanobacterial and algal cellulose synthase genes. Representative members of the CESA and CSL families from Arabidopsis, rice and *P. patens* and cyanobacterial, green algal and red algal cellulose synthases were used to search the *S. moellendorffii* genome with the internal tblastn function. All hits were examined and gene models were identified and edited manually.

The proposed gene names were used for the annotation in JGI's database and may be searched at <http://genome.jgi-psf.org/annotator/servlet/jgi.annotation.Annotation?pDb = Selmol>

Phylogenetic analysis

Sequences were aligned using ClustalX 2.0 with standard settings and applying the Gonnet substitution matrix series. Areas with poor resolution were deleted using Bioedit [111] and sequences were realigned. Several rounds of editing were done if needed. *P. patens*, rice or Arabidopsis sequences with obvious and large mistakes, be it annotation mistakes or pseudogenes, were not included in the trees, but protein ID are included in figure legends. Phylogenetic trees were generated by maximum likelihood using proML from the phylip package v3.68 [112] with 250 bootstraps from seqboot, using Jones-Taylor-Thornton matrix [113], no global rearrangement, rough analysis and one jumble. For clarification, cosmetic rearrangement of the trees was made using Adobe Illustrator (Adobe, USA).

Supporting Information

Figure S1 Phylogenetic tree of GT64. The family is not well characterized. Analysis of *epc1* (At3g55830) from clade A indicated that GT64 could be involved in cell wall biosynthesis. No putative function for the GT64A or the uncharacterized GT64B and C has been reported. (EPS)

Figure S2 Phylogenetic tree of GT48. This family includes the callose synthases abbreviated GSL for glucan synthase like. No other activities are known from GT48 and the *S. moellendorffii* members are distributed throughout the tree. GT48 candidates

that are not included due to incomplete sequences include At2g36850 and At5g36810. (EPS)

Figure S3 Phylogenetic tree of the CESA family. Arabidopsis genes required for synthesis of cellulose in primary (subgroups P1, P2, P3) and secondary (subgroups S1, S2, S3) cell walls have orthologs in rice and other seed plants. *S. moellendorffii* and *P. patens* CesAs form species-specific subgroups and lack orthologs of seed plant CesAs specialized for primary and secondary cell wall synthesis. The two *P. patens* genes marked with a * (PpCesA6 and PpCesA7) do not have JGI protein ID numbers because they failed to assemble completely due to their extremely high similarity. They are designated by the GenBank accession numbers for their full-length cDNA sequences. SmCESA5 is not included in the tree. (EPS)

Figure S4 Phylogenetic tree of the CslA family. Species-specific subgroups are well supported. Mannan synthase and glucomannan synthase activity has been demonstrated for proteins encoded by AtCslA1,2,3,7,9 and PpCslA1 (183385) and 2 (179490). GT2 CslC. A mixed Arabidopsis and rice subgroup and a rice subgroup are well supported. Heterologous expression of AtCslC4, a member of the mixed subgroup, resulted in production of b-1,4-glucan. Members not included in tree comprise Os07g0124750 and Os07g0630900. (EPS)

Figure S5 Phylogenetic tree of the CslD family. Arabidopsis genes with phenotypes in root hairs (AtCslD 2,3), pollen tubes (AtCslD 1,4), and stems (AtCslD5) all have rice orthologs. *S. moellendorffii* and *P. patens* CslDs form species-specific subgroups. (EPS)

Figure S6 Phylogenetic tree of GT34. Known activities, xyloglucan α -1,6-xylosyltransferases, AtXXT1 and AtXXT2 are found in clade A while AtXXT5 is in clade C. Arrows mark the likely Arabidopsis orthologs of the *Cyamopsis* galactomannan α -1,6-galactosyltransferase. The large B-clade comprises species-specific subgroups, but the basal topology is not resolved. (EPS)

Figure S7 Phylogenetic tree of GT47. The family includes inverting activities using a range of donors and acceptors. Naming of this family was established by Li et al., (2004). Subclade A contains xyloglucan galactosyltransferases (see also Figure 6), subclade B includes putative arabinan arabinosyltransferases, subclade C contains at least a xylogalacturonan xylosyltransferase (and probably other activities), subclade D contains members with unknown activity putatively involved in xylan biosynthesis (subgroup neighbouring to subclade E) and putatively involved in biosynthesis of the reducing end tetrasaccharide sequence of xylan (subgroup neighbouring to subclade F). Nothing is known about the activities present in subclade E or F. Candidates not included in tree comprise Os03g0182300, Os04g0633450, Os06g0176100, Pp156314, Pp201625, Pp214811, At4g13990, At5g37000. (EPS)

Figure S8 Phylogenetic tree of GT8. The GAUT and GATL clades of GT8 are presented here. The other members of GT8 are too divergent from the GAUT and GATL clades to be included in a single tree. The GAUT clade includes a homogalacturonan synthase (At3g61130, AtGAUT1) and an enzyme involved in homogalacturonan biosynthesis putatively a homogalacturonan synthase (At3g25140, AtQUA1). The GATL

clade is putatively involved in biosynthesis of the reducing end tetrasaccharide sequence of xylans based on knock out mutant analysis of PARVUS (At1g19300). Candidates not included in tree comprise Pp123164 and Os10g0454100. (EPS)

Figure S9 Phylogenetic tree of GT43. The family is divided into two subclades, both of which are likely to contain xylan backbone xylosyltransferases based on analysis of Arabidopsis mutants. (EPS)

Figure S10 Phylogenetic tree of the GT8 GUX clade. This clade has been shown to contain xylan glucuronosyltransferases. (EPS)

Figure S11 Phylogenetic tree of GT61. The family includes arabinoxylan arabinosyltransferases (Clade C). Beside this activity, AtXYLT, a β -1,2-xylosyltransferase involved in protein *N*-glycosylation is also found in GT61. It belongs to a clade of GT61 (Clade B), which is not presented in the tree due to low similarity to the A and C clades. Arrows indicate those among the rice sequences that in particular drive the grass bias of the C-clade in the Mitchell et al. (2007) study. Os06g0707000 is not included. (EPS)

Figure S12 Phylogenetic tree of GT77. Known activities include the RG-II α -1,3-xylosyltransferases RGXT1, RGXT2 and RGXT3. β -arabinosyltransferase activity is proposed for RRA1, RRA2 and Xeg113. Resolution in the part of the tree not assigned clade letter(s) is quite dependent on which species are included (data not shown). At5g40900 is in GT77 according to CAZy but has been excluded. (EPS)

Figure S13 Phylogenetic tree of GT10. The family includes fucosyltransferases involved in *N*-linked glycosylation of proteins. The FucTC clade includes α -1,4 fucosyltransferases involved in biosynthesis of the Lewis glycan structure. The FucTA and FucTB

clades include α -1,3 fucosyltransferases involved in biosynthesis of complex *N*-linked glycans. (EPS)

Figure S14 Phylogenetic tree of clade B of GT31. A galactosyltransferase activity involved in biosynthesis of the *N*-linked glucan (AtGalT1) is the only known activity. Two other subclades contain only flowering plant sequences. (EPS)

Figure S15 Phylogenetic tree of c-clade of GT31. The clade consists of two well resolved subclades with only flowering plant sequences and a subclade containing also *P. patens* and *S. moellendorffii* sequences. The basal topology is not resolved. No evidence for the function of GT31C has been published. At5g57500 is not included. (EPS)

Figure S16 Phylogenetic tree of the DUF266 family. This family is not presently recognized as a GT family but is likely to be included in CAZy as bioinformatic approaches and mutant studies have implicated the DUF266 proteins as GTs. Three clades are recognised. Members of the A clade have been proposed to be involved in AGP biosynthesis based on analysis of the *osbc10* mutant. At4g31350 is not included. (EPS)

Table S1 *S. moellendorffii* and *P. patens* protein IDs and gene names listed by family and clade. (XLS)

Author Contributions

Conceived and designed the experiments: PU JH HVS WGTW AR. Performed the experiments: PU JH AR IS JF. Analyzed the data: PU JH HVS WGTW AR IS JAB BLP JF. Contributed reagents/materials/analysis tools: WGTW IS AR HVS JAB JH PU JF. Wrote the paper: PU JH HVS WGTW AR IS JAB BLP JF. Background database: JAB.

References

- Banks JA (2009) Selaginella and 400 Million Years of Separation. *Annu. Rev. Plant Biol.* 60: 223–238.
- Wang W, Tanurdzie M, Luo M, Sisneros N, Kim HR et al (2005) Construction of a bacterial artificial chromosome library from the spikemoss *Selaginella moellendorffii*: a new resource for plant comparative genomics. *BMC Plant Biol.* 5: 10.
- Banks JO, Nishiyama T, Hasebe M, Bowman JL, Gribskov M et al (2011) The *Selaginella* Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science* 332: 960–963.
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.
- Xu Z, Zhang D, Hu J, Zhou X, Ye X, et al. (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics* 8:10 Suppl 11: S3.
- Popper ZA, Tuohy MG (2010) Beyond the Green: Understanding the Evolutionary Puzzle of Plant and Algal Cell Walls. *Plant Physiol* 153: 373–383.
- Sorensen I, Domozych D, Willats WGT (2010) How Have Plant Cell Walls Evolved? *Plant Physiol* 153: 366–372.
- Friedman WE, Cook ME (2000) The origin and early evolution of tracheids in vascular plants: integration of palaeobotanical and neobotanical data. *Phil. Trans. R. Soc. Lond. B* 355: 857–868.
- Raven JA (1993) The evolution of vascular plants in relation to quantitative functioning of dead water-conducting cells and stomata. *Biol. Rev.* 68: 337–363.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37(Database issue). pp D233–238.
- Moller I, Sorensen I, Bernal AJ, Blaukopf C, Lee K, Øbro J, et al. (2007) High-throughput mapping of cell wall polymers within and between plants using novel microarrays. *Plant J* 50: 1118–1128.
- Pena MJ, Darvill AG, Eberhard S, York WS, O'Neill MA (2008) Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* 18: 891–904.
- Popper ZA, Fry SC (2003) Primary cell wall composition of bryophytes and charophytes. *Ann Bot (Lond)* 91: 1–12.
- Liepman AH, Nairn CJ, Willats WGT, Sorensen I, Roberts AW, et al. (2007) Functional genomic analysis supports conservation of function among CslA gene family members and suggests diverse roles of mannan in plants. *Plant Physiol* 143: 1881–1893.
- Sorensen I, Pettolino FA, Wilson SM, Doblin MS, Johansen B, et al. (2008) Mixed-linkage (1 → 3), (1 → 4)-beta-D-glucan is not unique to the poales and is an abundant component of *Equisetum arvense* cell walls *Plant J* 54: 510–521.
- Tounou E, Takio S, Sakai A, Ono K, Takano H (2002) Ampicillin Inhibits Chloroplast Division in Cultured Cells of the Liverwort *Marchantia polymorpha*. *Cytologia* 67: 429–434.
- Machida M, Takechi K, Sato H, Chung SJ, Kuroiwa H, et al. (2006) Genes for the peptidoglycan pathway are essential for chloroplast division in moss. *Proc Natl Acad Sci USA* 103: 6753–6758.
- Izumi Y, Ono K, Takano H (2003) Inhibition of Plastid Division by Ampicillin in the Pteridophyte *Selaginella nipponica* Fr. et Sav. *Plant and Cell Physiology* 44: 181–188.
- Borges N, Ramos A, Raven NDH, Sharp RJ, Santos H (2002) Comparative study on the thermostabilising effect of mannosylglycerate and other compatible solutes on model enzymes. *Extremophiles* 6: 209–216.
- Singh SK, Eland C, Harholt J, Scheller HV, Marchant A (2005) Cell adhesion in *Arabidopsis thaliana* is mediated by ECTOPICALLY PARTING CELLS 1 – a glycosyltransferase (GT64) related to the animal exostosins. *Plant J* 43: 384–397.
- Bown L, Kusaba S, Goubet F, Codrai L, Dale AG, et al. (2007) The ectopically parting cells 1–2 (*epc1–2*) mutant exhibits an exaggerated response to abscisic acid. *J Exp Bot* 58: 1813–1823.
- Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. *Plant Physiol* 124: 495–498.

23. Delmer DP (1999) Cellulose biosynthesis: Exciting times for a difficult field of study. *Annu Rev Plant Physiol. Plant Mol Biol* 50: 245–276.
24. Stone BA, Jacobs AK, Hrmova M, Burton RA, Fincher GB (2010) Biosynthesis of Plant Cell Wall and Related Polysaccharides by Enzymes of the GT2 and GT48 Families. In: *Plant Polysaccharides* (Ulvskov ed.): 09–153, Wiley-Blackwell.
25. Saxena I, Brown R (2000) Cellulose synthases and related enzymes *Current Opin Plant Biol* 3: 523–531.
26. Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proc Natl Acad Sci USA* 100: 1450–1455.
27. Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, et al. (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci Unit States Am* 104: 15572–15577.
28. Persson S, Paredes A, Carroll A, Palsdottir H, Doblin M, et al. (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *Proc Natl Acad Sci Unit States Am* 104: 15566–15571.
29. Nairn CJ, Haselkorn T (2005) Three loblolly pine CesA genes expressed in developing xylem are orthologous to secondary cell wall CesA genes of angiosperms. *New Phytol* 166: 907–915.
30. Kumar M, Thammannagowda S, Bulone V, Chiang V, Han KH, et al. (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. *Trends Plant Sci* 14: 248–254.
31. Roberts AW, Bushoven JT (2007) The cellulose synthase (CESA) gene superfamily of the moss *Physcomitrella patens*. *Plant Mol. Biol.* 63: 207–219.
32. Weng JK, Li X, Stout J, Chapple C (2008) Independent origins of syringyl lignin in vascular plants. *Proc Natl Acad Sci U S A*. 105: 7887–7892.
33. Friedman WE, Cook ME (2000) The origin and early evolution of tracheids in vascular plants: integration of palaeobotanical and neobotanical data. *Phil. Trans. R. Soc. Lond. B* 355: 857–868.
34. Nobles DR, Romanovitch DK, Brown RM (2001) Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? *Plant Physiol* 127: 529–542.
35. Roberts EM, Roberts AW (2009) A cellulose synthase (CESA) gene from the red alga *Porphyra yezoensis* (Rhodophyta). *J. Phycol.* 45: 203–212.
36. Roberts AW, Roberts EM (2007) Evolution of the cellulose synthase (CESA) gene family: insights from green algae and seedless plants. In: *Cellulose: Molecular and Structural Biology*, R. M. Brown, Jr. and I. M. Saxena, eds. Springer, Berlin, 17–34.
37. Tsekos I, Reiss H-D (1994) Tip growth and the frequency and distribution of cellulose microfibril-synthesizing complexes in the plasma membrane of apical shoot cells of the red alga *Porphyra yezoensis*. *J Phycol* 30: 300–310.
38. Dhugga KS (2001) Building the wall: genes and enzyme complexes for polysaccharide synthases. *Curr Opin Plant Biol* 4: 488–493.
39. Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, et al. (2007) A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proc Natl Acad Sci Unit States Am* 104: 8550–8555.
40. Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, et al. (2006) Cellulose synthase-like CSLF genes mediate the synthesis of cell wall (1,3;1,4)- β -D-glucans. *Science* 311: 1940–1942.
41. Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, et al. (2009) A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic *Arabidopsis*. *Proc Natl Acad Sci Unit States Am* 106: 5996–6001.
42. Scheller HV, Ulvskov P (2010) Hemicelluloses. *Annu Rev Plant Biol* 61: 263–289.
43. Yin YB, Huang JL, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biology* 9: 99.
44. Yin L, Verhertbruggen Y, Oikawa A, Manisseri C, Knierim B, et al. (2011) The Cooperative Activities of CSLD2, CSLD3, and CSLD5 Are Required for Normal *Arabidopsis* Development. *MPlant* doi: 10.1093/mp/ssr026.
45. Bernal AJ, Jensen JK, Harholt J, Sorensen S, Moller I, et al. (2007) Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in *Arabidopsis*. *Plant J.* 52: 791–802.
46. Manfield IW, Orfila C, McCartney L, Harholt J, Bernal AJ, et al. (2004) Novel cell wall architecture of isoxaben-habituated *Arabidopsis* suspension-cultured cells: global transcript profiling and cellular analysis *Plant J.* 40: 260–275.
47. Park S, Szumlanski AL, Gu F, Guo F, Nielsen E (2011) A Role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells. *Nat. Cell Bio.* 13: 973–980.
48. Cavalier DM, Lerouxel O, Neumetzler L, Yamauchi K, Reinecke A, et al. (2008) Disrupting two *Arabidopsis thaliana* xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. *Plant Cell* 20: 1519–1537.
49. Zabolina OA, van de Ven WTG, Freshour G, Drakakaki G, Cavalier D, et al. (2008) The *Arabidopsis* XT5 protein encodes a putative α -1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. *Plant J.* 56: 101–115.
50. Pettolino F, Hoogenraad NJ, Ferguson C, Bacic A, Johnson E, et al. (2001) A (1-4)- β -mannan-specific monoclonal antibody and its use in the immunocytochemical location of galactomannans. *Planta* 214: 235–242.
51. Ordaz-Ortiz JJ, Marcus SE, Knox JP (2009) Cell wall microstructure analysis implicates hemicellulose polysaccharides in cell adhesion in tomato fruit pericarp parenchyma. *Mol Plant* 2: 910–21.
52. Madson M, Dunand C, Li X, Verma R, Vanzin GF, et al. (2003) The MUR3 gene of *Arabidopsis* encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins *Plant Cell* 15: 1662–70.
53. Li XM, Cordero I, Caplan J, Molhoj M, Reiter WD (2004) Molecular analysis of 10 coding regions from *Arabidopsis* that are homologous to the MUR3 xyloglucan galactosyltransferase. *Plant Physiol* 134: 940–950.
54. Sarria R, Wagner TA, O'Neill MA, Faik A, Wilkerson CG, et al. (2001) Characterization of a Family of *Arabidopsis* Genes Related to Xyloglucan Fucosyltransferase. *Plant Physiol* 127: 1595–1606.
55. Wu YY, Williams M, Bernard S, Driouch A, Showalter AM, et al. (2010) Functional Identification of Two Nonredundant *Arabidopsis* alpha(1,2)Fucosyltransferases Specific to Arabinogalactan Proteins. *J Biol Chem* 285: 13638–13645.
56. Burke D, Kaufman P, McNeil M, Albersheim P (1974) The Structure of Plant Cell Walls VI. A Survey of the Walls of Suspension-cultured Monocots *Plant Physiol* 54: 109–115.
57. Labavitch JM, Ray PM (1978) Structure of Hemicellulosic Polysaccharides of *Avena sativa* Coleoptile Cell-Walls *Phytochem* 17: 933–93.
58. Kato Y, Ito S, Iki K, Matsuda K (1982) Xyloglucan and beta-d-glucan in cell-walls of rice seedlings. *Plant Cell Physiol* 23: 351–364.
59. Yokoyama R, Nishitani K (2004) Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and *Arabidopsis*. *Plant Cell Physiol* 45: 1111–1121.
60. Hayashi T (1989) Xyloglucans in the Primary Cell Wall. *Annu Rev Plant Physiol Plant Mol Biol* 40: 139–168.
61. McDougall GJ, Fry SC (1994) Fucosylated Xyloglucan in Suspension-Cultured Cells of the Gramineous Monocotyledon, *Festuca-Arundinacea*. *J Plant Physiol* 143: 591–595.
62. Günl M, Pauly M (2010) AX3Y3 encodes a α -xylosidase that impacts the structure and accessibility of the hemicellulose xyloglucan in *Arabidopsis* plant cell walls. *Planta* 233: 707–19.
63. York WS, O'Neill MA (2008) Biochemical control of xylan biosynthesis — which end is up? *Curr Opin Plant Biol* 11: 258–265.
64. Lao NT, Long D, Kiang S, Coupland G, Shoue DA, et al. (2003) Mutation of a family 8 glycosyltransferase gene alters cell wall carbohydrate composition and causes a humidity-sensitive semi-sterile dwarf phenotype in *Arabidopsis*. *Plant Mol Biol* 53: 647–661.
65. Wu AM, Hornblad E, Voxeur A, Gerber L, Rihouey C, et al. (2010) Analysis of the *Arabidopsis* IRX9/IRX9-L and IRX14/IRX14-L Pairs of Glycosyltransferase Genes Reveals Critical Contributions to Biosynthesis of the Hemicellulose Glucuronoxylan. *Plant Physiol* 153: 542–554.
66. Brown DM, Zhang ZN, Stephens E, Dupree P, Turner SR (2009) Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in *Arabidopsis*. *Plant J* 57: 732–746.
67. Wu AM, Rihouey C, Seveno M, Hornblad E, Singh SK, et al. (2009) The *Arabidopsis* IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. *Plant J* 57: 718–731.
68. Mutwil M, Obro J, Willats WG, Persson S (2008) GeneCAT – novel webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res.* 1; 36(Web Server issue). pp W320–6.
69. Mortimer JC, Miles GP, Brown DM, Zhang ZN, Segura MP, et al. (2010) Absence of branches from xylan in *Arabidopsis* gux mutants reveals potential for simplification of lignocellulosic biomass. *Proc Natl Acad Sci U S A* 107: 17409–17414.
70. Oikawa A, Joshi HJ, Rennie EA, Ebert B, Manisseri C, et al. (2010) An Integrative Approach to the Identification of *Arabidopsis* and Rice Genes Involved in Xylan and Secondary Wall Development. *Plos One* 5.
71. Chatterjee M, Berbezzy P, Vyas D, Coates S, Barsby T (2005) Reduced expression of a protein homologous to glycogenin leads to reduction of starch content in *Arabidopsis* leaves. *Plant Science* 168: 501–509.
72. Mitchell RA, Dupree P, Shewry PR (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiol* 144: 43–53.
73. Strasser R, Mucha J, Mach L, Altmann F, Wilson IBH, et al. (2000) Molecular cloning and functional expression of beta 1,2-xylosyltransferase cDNA from *Arabidopsis thaliana*. *FEBS Lett* 472: 105–108.
74. Wende G, Fry SC (1997) 2-O-beta-D-xylopyranosyl-(5-O-feruloyl)-L-arabinose, a widespread component of grass cell walls. *Phytochemistry* 44: 1019–1030.
75. Anders N, Wilkinson MD, Lovegrove A, Freeman J, Tryfona T, et al. (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proc Natl Acad Sci U S A.* 17: 989–93.
76. Mohnen D (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol* 11: 266–277.
77. Scheller H, Jensen JK, Sorensen SO, Harholt J, Geshi N (2007) Biosynthesis of pectin. *Physiol Plantarum* 129: 283–295.
78. Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res* 344: 1879–1900.
79. Harholt J, Suttangkakul A, Scheller HV (2010) Biosynthesis of Pectin. *Plant Physiol* 153: 384–395.



80. Liepman AH, Wightman R, Geshi N, Turner SR, Scheller HV (2010) Arabidopsis – a powerful model system for plant cell wall research. *Plant J* 61: 1107–21.
81. Yin YB, Chen HL, Hahn MG, Mohnen D, Xu Y (2010) Evolution and Function of the Plant Cell Wall Synthesis-Related Glycosyltransferase Family 8. *Plant Physiol* 153: 1729–1746.
82. Sterling JD, Atmodjo MA, Inwood SE, Kumar Kolli VS, Quigley HF, et al. (2006) Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proc Natl Acad Sci Unit States Am* 103: 5236–5241.
83. Bouton S, Leboeuf E, Mouille G, Leydecker MT, Talbotec J et al (2002) QUASIMODO1 encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in Arabidopsis. *Plant Cell* 14: 2577–2590.
84. Orfila C, Sorensen SO, Harholt J, Geshi N, Crombie H et al (2004) QUASIMODO1 is expressed in vascular tissue of Arabidopsis thaliana inflorescence stems, and affects homogalacturonan and xylan biosynthesis. *Planta* 222: 613–622.
85. Harholt J, Jensen JK, Sorensen SO, Orfila C, Pauly M, et al. (2006) ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in Arabidopsis. *Plant Physiol* 140: 49–58.
86. Jensen JK, Sorensen SO, Harholt J, Geshi N, Sakuragi Y, et al. (2008) Identification of a xylogalacturonan xylosyltransferase involved in pectin biosynthesis in Arabidopsis. *Plant Cell* 20: 1289–302.
87. Egelund J, Petersen BL, Motawia MS, Damager I, Faik A, et al. (2006) Arabidopsis thaliana RGXT1 and RGXT2 encode Golgi-localized (1,3)-alpha-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. *Plant Cell* 18: 2593–2607.
88. Egelund J, Damager I, Faber K, Ulvskov P, Petersen BL (2008) Functional characterisation of a putative Rhamnogalacturonan II specific XylosylTransferase. *FEBS Lett* 582: 3217–3222.
89. Matsunaga T, Ishii T, Matsumoto S, Higuchi M, Darvill A, et al. (2004) Occurrence of the Primary Cell Wall Polysaccharide Rhamnogalacturonan II in Pteridophytes, Lycophytes, and Bryophytes. Implications for the Evolution of Vascular Plants. *Plant Physiol* 134: 339–351.
90. Xu JF, Tan L, Lampion DTA, Showalter AM, Kieliszewski MJ (2008) The O-Hyp glycosylation code in tobacco and Arabidopsis and a proposed role of Hyp-glycans in secretion. *Phytochem* 69: 1631–1640.
91. Egelund J, Obel N, Ulvskov P, Geshi N, Pauly M, et al. (2007) Molecular characterization of two Arabidopsis thaliana glycosyltransferase mutants, rra-1 and -2, which have a reduced content of arabinose in a polymer tightly associated with the cellulose residue. *Plant Mol Biol* 64: 439–451.
92. Gille S, Hänsel U, Ziemann M, Pauly M (2009) Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proc Natl Acad Sci Unit States Am* doi: 10.1073/pnas.0905434106.
93. Velasquez SM, Ricardi MM, Dorosz JG, Fernandez PV, Nadra AD, et al. (2011) O-glycosylated cell wall proteins are essential in root hair growth. *Science* 17: 1401–3.
94. Bollig K, Lamshoeft M, Schweirner K, Marner FJ, Budzikiewicz H, et al. (2007) Structural analysis of linear hydroxyproline-bound O-glycans of Chlamydomonas reinhardtii-conservation of the inner core in Chlamydomonas and land plants. *Carbohydr Res* 342: 2557–2566.
95. Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. *Ann Rev Plant Biol* 58: 137–161.
96. Qu YM, Egelund J, Gilson PR, Houghton F, Gleeson PA, et al. (2008) Identification of a novel group of putative Arabidopsis thaliana α -(1,3)-galactosyltransferases. *Plant Mol Biol* 68: 43–59.
97. Strasser R, Singh Bondili J, Vavra U, Schoberer J, et al. (2007) A Unique beta-1,3-Galactosyltransferase Is Indispensable for the Biosynthesis of N-Glycans Containing Lewis a Structures in Arabidopsis thaliana. *Plant Cell* 19: 2278–2292.
98. Narimatsu H (2006) Human glycogene cloning: focus on beta 3-glycosyltransferase and beta 4-glycosyltransferase families. *Curr Opin Struct Biol* 16: 567–75.
99. Egelund J, Skjoldt M, Geshi N, Ulvskov P, Petersen BL (2004) A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes. *Plant Physiol* 136: 2609–2620.
100. Sadowski PG, Groen AJ, Dupree P, Lilley KS (2008) Sub-cellular localization of membrane proteins. *Proteomics* 8: 3991–4011.
101. Hansen SF, Bettler E, Wimmerova M, 4, Imberty A, Lerouxel O, et al. (2009) Combination of Several Bioinformatics Approaches for the Identification of New Putative Glycosyltransferases in Arabidopsis. *Journal of Proteome Research* 8 (Sp. Iss. SI). pp 743–753.
102. Zhou YH, Li SB, Qian Q, Zeng DL, Zhang M, et al. (2009) BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (*Oryza sativa* L.). *Plant J* 57: 446–462.
103. Yin YB, Huang JL, Xu Y (2008) The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biology* 9: 99.
104. Clausen MH, Willats WGT, Knox JP (2003) Synthetic methyl hexagalacturonate hapten inhibitors of antihomogalacturonan monoclonal antibodies LM7, JIM5 and JIM7. *Carbohydr Res* 338: 1797–1800.
105. McCartney L, Marcus SE, Knox JP (2005) Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *Journal of Histochemistry and Cytochemistry* 53: 543–546.
106. Jones L, Seymour GB, Knox JP (1997) Localization of pectic galactan in tomato cell walls using a monoclonal antibody specific to (1->4)-beta-D-galactan. *Plant Physiol* 113: 1405–1412.
107. Willats WGT, Marcus SE, Knox JP (1998) Generation of a monoclonal antibody specific to (1->5)-alpha-L-arabinan. *Carbohydr Res* 308: 149–152.
108. Obro J, Harholt J, Scheller HV, Orfila C (2004) Rhamnogalacturonan I in Solanum tuberosum tubers contains complex arabinogalactan structures. *Phytochemistry* 65: 1429–1438.
109. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, et al. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 35: D237–40.
110. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat Protocol* 4: 363–371.
111. Hall A (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95–98.
112. Felsenstein, J (1989) PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164–166.
113. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 8: 275–282.